The Thesis committee for Michael Eric Ferguson

Certifies that this is the approved version of the following thesis:

High-accuracy *ab initio* thermochemistry: Application to hydrocarbons

APPROVED BY

SUPERVISING COMMITTEE:

Supervisor: _____
John F. Stanton

_____
Eric V. Anslyn

High-accuracy *ab initio* thermochemistry: Application to hydrocarbons

by

Michael Eric Ferguson, B.S.

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF ARTS

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2013

Dedicated to Edgar, who supported me always.

Acknowledgments

I am deeply in debt to Dr. John F. Stanton for his advice and support. Without his guidance, I would never have been able to complete my degree. Without the work of Branko Ruscic on the ATcT thermochemical tables to compare our results with, this work would not have been possible. Furthermore, I would like to thank Dr. Eric Anslyn for agreeing to be on my thesis committee despite his busy schedule.

Finally, I would like to thank all of my friends and family who supported me along the way. Their love, support, and encouragement allowed me to make it as far as I have in life. I will be forever grateful to them for all they have done for me.

High-accuracy *ab initio* thermochemistry: Application to hydrocarbons

Michael Eric Ferguson, M.A.
The University of Texas at Austin, 2013

Supervisor: John F. Stanton

This work focuses on an examination of the high-accuracy extrapoloated *ab initio* thermochemistry (HEAT) protocol of determining molecular atomization energies. The HEAT protocol does not utilize experimental data or empirical scaling effects. The accuracy of the approach is tested via comparison to ATcT data, and all molecules fall within 1 kcal mol^{-1} of accepted values. There are several important points to note about this treatment: namely, that we have used atomic natural orbital (ANO) basis sets for the calculation of the zero point energy and that we have made determinations for larger molecules than previously done with HEAT. The molecules in this paper were chosen to provide benchmark numbers for the homodesmotic reaction heirarchy as described by Wheeler et al.[6] The relative accuracy of the approach is considered, as well as a discussion of possible remaining sources of error.

Table of Contents

Chapter 1

Introduction

1.1 Purpose/Rationale

Due to advances in both hardware and theoretical methods, it is now possible to calculate with great accuracy thermochemical data for many molecules. It is relevant to note that the available theoretical accuracy is *greater* than the standard of "chemical accuracy" (<1 kcal mol^{-1}).[1–3,17,25–27] This possibility grants researchers the ability to use theoretical methods to determine thermochemical data of interest, principally that of bond energies and heats of formation. Other methods based upon focal-point extrapolation[24] or the Weizmann (Wn) family of methods[25–27] have also recently gained similar advances towards an accuracy of <1 kJ mol^{-1}, significantly more than chemical accuracy (4.184 kcal = 1 kJ).

The purpose of this work is to study new approximations in the high-accuracy extrapolated *ab initio* thermochemistry (HEAT) protocol[1–3], specifically, to the anharmonicity of the zero-point energy and its application to larger molecules. The main advantage of the HEAT protocol is that it is a more rigorous approach than many competing methods. Therefore, while many researchers have moved to less accurate but less costly density func-

tional methods, wave-function methods can provide very accurate measures of fundamental vibrational frequencies[28–30] (within $\approx 10\,cm^{-1}$), bond lengths[7] (within $\approx$ 0.002-0.003 Å), and bond angles[7] (within $\approx 0.5^{\circ}$). Additionally, as mentioned above and principal to this piece of work, research methods now allow thermodynamic values such as enthalpies of formation, heat capacities, and standard entropies to be determined to within chemical accuracy.

The level at which modern *ab initio* methods can be competitive with experimental values depends largely upon the size of the molecule in question. It is obvious that for relatively small molecules that theory can provide some data equal to, or arguably better than, experimental values. Although HEAT is limited, like other methods, by the size of the molecules (indeed the largest molecule attempted in this paper, neopentane, has only 17 atoms), this approach can give valuable information for molecules that are difficult to study experimentally. Specifically, transient species and radicals in atmospheric and combustion reactions, which require highly accurate thermochemical data, are notoriously difficult to measure in lab. Therefore, we believe that the development of high-accuracy *ab initio* thermochemical methods are not an academic exercise with no practical benefit, but that these efforts will allow increasingly accurate data for larger and larger molecules as a result of the advance of hardware and theoretical models.

The main advantage of HEAT, therefore, is its high level of accuracy. The main quantities we are concerned with are not total energies of molecules, but rather energy differences. Therefore, when calculations are performed at

the same level of theory, errors in a given contribution to the total energy likely originate from the same source, and thus will cancel when the difference is taken.[1] It is also worthwhile to note that the "accuracy" in chemistry is often represented with a 95% confidence limit, rather than a standard deviation. We believe this distinction is important because the definition of chemical accuracy should be within the same 95% confidence limit. This standard was proposed by Rossini[31] and has since been adopted by almost all other thermochemical databases such as CODATA,[32] Janaf,[33] and Gurvich *et al.*[34] Therefore, other methods that do not require this distinction, especially that of the mean absolute deviation method of reporting error, drastically underestimate the error of the approach.[3]

Also of great importance to the development of the HEAT protocol is the improvement of the precision of existing data due to the active table approach of Ruscic and co-workers.[23] The reduction of the uncertainty in the experimental values for key species, specifically the carbon atom, has helped reduce the amount of uncertainty in the resulting heats of formation from theory. We should note however, that $\Delta_f H^\circ$ for carbon still has a significant amount of error associated with it (0.06 kJ mol^{-1}). Although this error may be small for most applications, when performing high-accuracy calculations such as those here, the cumulative effect of this error can be significant. Therefore a systematic error exists for all organic molecules using the atomization energy approach, proportional to the number of carbon atoms.

The primary goal of this paper, however, is to establish benchmark

energies for the homodesmotic reaction hierarchy as described by Wheeler et al.[6] which establishes a reaction hierarchy to describe the similarity of products and reactants with respect to their chemical environment. The 11 species selected were chosen due to their inclusion as reactants studied in the above paper. With the establishment of benchmark $\Delta_f H^\circ$ for these species, we hope that it will be possible to apply the results of this paper to other chemically relevant reactions using the isodesmic reaction format described below.

1.2 Enthalpy and heats of formation

Enthalpy is a state function that, like energy, it is always conserved and has no natural zero. Enthalpy differences come from an arbitrarily defined zero, or standard state. Experimental standard enthalpies of formation refer to the energy difference between a molecule and its constituent atoms in their standard states.

$$\mathrm{C(s,graphite)} + 2\,\mathrm{H_2(g)} \longrightarrow \mathrm{CH_4(g)} \tag{1.1}$$

The difference in the enthalpy of the molecule and in its standard state and the constituent elements in their standard states is known as the enthalpy of formation.

A second reference state used in chemistry is that of the free nuclei and electrons in the gaseous state which combine to form the molecule. The energy difference between the molecule and its separated atoms is referred to

as the total molecular energy.

$$C^{6+} + 4\,H^{+} + 10\,e^{-} \longrightarrow CH_4(g) \tag{1.2}$$

Therefore, two common points of reference in chemistry are different and provide different zeroes for the energy. Another common energy difference used in chemistry is the atomization energy, referring to the enthalpy change when a compound's bonds are broken and the component atoms are reduced to individual atoms.

$$CH_4(g) \longrightarrow C(g) + 4\,H(g) \tag{1.3}$$

Therefore, atomization energies differ from enthalpies of formation in that they are always positive (because all bonds are broken and none are formed) and the products are always monatomic.

The most prevalent method of determing heats of formation of molecules using model chemistry utilizes atomization energies as the basis for thermochemistry. In other words, the total molecule energies are calculated at some level of theory, and the atomization energies are then calculated using total atomic energies obtained at the same level of theory. For example, for the molecule CH_4,

$$AE(CH_4(g)) = E(C(g)) + 4\,E(H(g)) - E(CH_4(g)) \tag{1.4}$$

where $AE(CH_4(g))$ is the atomization energy of $CH_4(g)$. We can determine the enthalpy of formation of CH_4 at 0K, for example, with Hess's law by

$$\Delta_f H^{\circ}(CH_4) = \Delta_f H^{\circ}(C(g)) + 4\,\Delta_f H^{\circ}(H(g)) - AE(CH_4(g)) \tag{1.5}$$

where $\Delta_f H^\circ$ values are taken from experimental data.

The reference state favored by HEAT is a reaction where the chemical environment of as many atoms as possible is preserved, but still permitting a reaction. This type of reaction is often called "isodesmic". However, the term isodesmic can be confusing in that it can refer to a whole range of reaction types. The distinction between each type of reaction and the varying degrees of accuracy are explored more by Wheeler et al.[6]. For the purposes of this paper, we will refer to a chemical reaction in which the relative chemical environment of both products and reactants is preserved as an isodesmic reaciton. The reaction

$$CH_3^- + CH_3X \rightarrow CH_4 + CH_2X^-$$

(where X = F, Cl, Br, I) is isodesmic because both before and after the reaction there are 6 C-H single bonds and one C-X single bond. The determination of heats of formation using isodesmic processes are known to have the highest accuracy.[4–6] The ideal situation is when an isodesmic process can be constructed in which *all* participants, other than the reaction molecule of interest, have precise, experimentally-measured enthalpies of formation. Thus, the calculation of precise $\Delta_f H^\circ$ for the species in this study will allow formulation of isodesmic reactions in which more of the participants have precisely known benchmark values, which is the major goal of this project.

The original HEAT protocol called for a model chemistry using reaction-based approaches for the molecule of interest. The primary results of this paper will be the atomization energies of the species studied, due to the convenient

ability to compare accuracy of the results using the method described above with ATcT.[23] Our primary goals are: to examine the accuracy in heats of formation for molecules larger than those originally possible when the HEAT protocol was first devised and to provide benchmark values for further analysis using the isodesmic reaction scheme.

1.3 Introduction to HEAT protocol

The HEAT protocol, as discussed above, is a model chemistry aimed at determining accurate thermochemical parameters. In order to determine total atomization energies and heats of formation, it is necessary to calculate the ground state energy of the target molecule or atom. Within the Born-Oppenheimer approximation, we can separate the energy contributions into vibrational and electronic components. The vibrational and electronic energies are given by the lowest eigenvalues of: the nuclear Hamiltonian and the electronic Hamiltonian (at equilibrium geometry).

The vibrational contribution is given by the expectation value of the nuclear kinetic energy operator over the electronic wavefunction. Calculation of the electronic contribution to the ground state energy involves the nonrelativistic electronic Hamiltonian. Afterwards, we can improve on this nonrelativistic, post Born-Oppenheimer picture and include scalar relativistic effects by using perterbation theory, the lowering of the ground state energy by spin-orbit interactions, and the diagonal Born-Oppenheimer correction to account for the error in the Born-Oppenheimer approximation.

In practice, all of these calculations are, by necessity, approximations. In order for the results to be exact, infinite basis sets would have to be used, which is impossible. Furthermore, full configuration interaction (FCI) is only possible (for modern computers) with very small molecules. Even if we were able to obtain FCI for a given molecule, the resulting energies would contain some error due to the use of one-particle basis sets.

In the HEAT protocol[1], the total ground state energy of an atom or molecule is given by

$$\mathrm{E} = \mathrm{E}^{\infty}_{\mathrm{HF}} + \Delta\mathrm{E}^{\infty}_{\mathrm{CCSD(T)}} + \Delta\mathrm{E}_{\mathrm{CCSDT}} + \Delta\mathrm{E}_{\mathrm{CCSDTQ}} + \Delta\mathrm{E}_{\mathrm{rel}} + \Delta\mathrm{E}_{\mathrm{SO}} + \Delta\mathrm{E}_{\mathrm{DBOC}} + \Delta\mathrm{E}_{\mathrm{ZPE}} \quad (1.6)$$

where $\mathrm{E}^{\infty}_{\mathrm{HF}}$ is the Hartree-Fock limit, $\Delta E^{\infty}_{CCSD(T)}$ is the CCSD(T) correlation energy,[12] ΔE_{CCSDT} is the estimated error in changing from CCSD(T) to the full triples calculation,[13] ΔE_{CCSDTQ} is the energy change due to higher level correlation effects (quadruple excitations),[14,15] ΔE_{rel} accounts for scalar relativistic effects as approximated by the two-electron Darwin term, ΔE_{SO} represents the difference between the true ground-state energy and the weighted average of spin-orbit levels, ΔE_{DBOC} is the diagonal Born-Oppenheimer correction, and ΔE_{ZPE} is the vibrational zero point energy.

1.3.1 Molecular Geometry

The molecular geometries used were obtained from geometry optimizations performed at the CCSD(T) level of theory with the cc-pVQZ basis sets.

The use of this geometry is justified by a study that suggests that at this level of theory, equilibrium geometry can be obtained to within <0.003 Å and <0.5° for bonds and bond angles, respectively.[7] The original HEAT protocol calls for the use of UHF for open-shell molecules. The error associated with this approach is addressed in the original HEAT paper[1] and is not relevent in this study. The RHF method was used for all molecules in this study, because all are closed-shell.

1.3.2 SCF and electron correlation

The HF-SCF energy and electron correlation in molecules constitute the bulk of the total energy that we calculate. It should be noted that the HEAT protocol does not distinguish between core and valence electrons. Rather, all electrons are treated equally using the aug-cc-pCVXZ series of basis sets. For the total Hartree-Fock energy within the Born-Oppenheimer approximation, the infinite basis set limit was extrapolated[16] using

$$E_{HF}^{X} = E_{HF}^{\infty} + a\exp(-bX) \tag{1.7}$$

where a and b are parameters and E_{HF}^{X} is the HF-SCF energy for a given aug-cc-pCVXZ basis set, where X = T(3), Q(4), and 5.[9–11] With three calculations for X = 3,4,5, it is possible to solve for the infinite basis set limit E_{HF}^{∞}. In Ref 3, other schemes were considered, such as an extrapolation based upon X = 4, 5, 6.[3] From these results, we can conclude that the most effective compromise between accuracy and computation time comes with the HEAT-345(Q) based method.

For electron correlation within the CCSD(T) method, the following formula was used to find the CCSD(T) correlation energy[18,19]:

$$\Delta E^{X}_{CCSD(T)} = \Delta E^{\infty}_{CCSD(T)} + \frac{a}{X^3} \tag{1.8}$$

where $\Delta E^{X}_{CCSD(T)}$ is the correlation energy, $\Delta E^{\infty}_{CCSD(T)}$ is that energy extrapolated to the infinite basis set limit. For the CCSD(T) correlation energy, X = 4,5 was used. It should be noted here that the correlation energy is not the total CCSD(T) energy, which includes the HF-SCF energy, but rather the difference between the total CCSD(T) energy and the total HF-SCF energy.

Although CCSD(T) has become the "gold standard" for computational chemistry, in cases where triples corrections are large or when there are severe problems with the reference function, it can be necessary to compare the energy difference between CCDST and CCSD(T) calculations. For full triples calculations, smaller basis sets are used in order to bring computation time down to a realistic level on modern hardware. Therefore,

$$\Delta E_{CCSDT} = E_{CCSDT}(fc) - E_{CCSD(T)}(fc) \tag{1.9}$$

where (fc) denotes frozen core electrons. All energies were obtained using CCSDT and CCSD(T) with the cc-pVTZ and cc-pVQZ basis sets. Each term in the above equation was obtained using the correlation energy extrapolation formula, Equation 1.8.

In order to obtain the high-accuracy results that are the goal of the HEAT protocol, quadruple excitations must also be considered.[35–39] A similar

reduction in basis set size accompanies the transition to quadruple excitations. Therefore, the equation to give the CCSDTQ correlation energy is given by

$$\Delta E_{CCSDTQ} = E_{CCSDTQ}(fc) - E_{CCSDT}(fc) \tag{1.10}$$

with the cc-pVDZ basis set in the frozen core approximation. Until recently, limitations in coupled-cluster code prohibited CCSDTQ calculations on larger molecules. With the development of general coupled-cluster codes,[40–45] we can now calculate quadruple excitation energies for relatively large molecules (for example neopentane with 17 atoms and 42 electrons). For the CCSDTQ calculations in this paper, the coupled cluster program MRCC[8] coupled to CFOUR was used.

1.3.3 Zero-point vibrational Energy

The zero-point vibrational energy levels described by spectroscopic rovibrational perterbation theory[22] are given by the equation

$$E(v) = G_0 + \sum_i \omega_i (v_i + \frac{1}{2}) + \sum_{i \geq j} x_{ij} (v_i + \frac{1}{2})(v_j + \frac{1}{2}) \tag{1.11}$$

where ω_i are the harmonic frequencies and x_{ij} are the anharmonicity constants. It follows, therefore, that the ground state (zero-point) energy is given by

$$E_{ZPE} = G_0 + \sum_i \frac{\omega_i}{2} + \sum_{i<j} \frac{x_{ij}}{4} \tag{1.12}$$

In the original HEAT protocol, the G_0 term was neglected due to lack of adequate theoretical models. Since that time, formulas for the G_0 term have

been derived[47–49] and we include G_0 in all ZPE calculations in this paper. Furthermore, in this work, the anharmonic portion of the ZPE is determined using the atomic natural orbital (ANO) basis set of Almlöf and Taylor.[20,21] The ANO0 basis set, roughly analagous to the cc-pVDZ basis set, was used instead of the cc-pVQZ basis set advocated in the first HEAT paper. We rationalize this reduction in basis set size, allowing the calculation of the anharmonic contribution for the larger molecules in this paper, per a recent study showing that the ANO basis sets perform well for ZPE calculations.[63,64]

1.3.4 Diagonal Born-Oppenheimer correction

The diagonal Born-Oppenheimer correction (DBOC) can be viewed as the first-order correction to the electronic energy associated with the nuclear kinetic energy operator.[3] This correction attempts to account for the general Born-Oppenheimer approximation that nuclear motion is negligible when solving the electronic wavefunction. The diagonal Born-Oppenheimer correction is given by

$$\Delta E_{DBOC} = \langle \Psi_e(r;R) | \hat{T}_n | \Psi_e(r;R) \rangle \tag{1.13}$$

where $\hat{T}_n$ is the nuclear kinetic energy operator, r is a set of electronic coordinates, R is a set of nuclear coordinates and $\Psi_e(r;R)$ is the normalized electronic wave function.[50–57]

In Ref. 3, the authors explored the advantages of a coupled cluster approach for the DBOC at the CCSD level of theory. It was found that the DBOC contribution changes little from going from HF-SCF to the CCSD level

of theory (the HF-SCF is expected to be accurate to within 0.1 kJ mol^{-1}).[3] In this work, this correction is calculated at the SCF level of theory with the aug-cc-pVTZ basis set.

1.3.5 Scalar relativistic effects

When considering relativity, the mass of the electron increases as a function of its velocity. This effect leads to orbital contraction and will slightly alter energy values. Therefore, the term ΔE_{rel} refers to this difference in energy and is included as a post-Hartree Fock correction. In the original HEAT protocol, relativistic corrections were obtained via perturbation theory based on the one-electron mass-velocity and Darwin terms. In this work, we use the expanded treatment explained in Ref. 2.[2,59,60] Specifically, the treatment is expanded to include the corresponding two-electron Darwin term. We performed all calculations at the CCSD(T) level of theory with aug-cc-pCVTZ basis.

1.3.6 Spin-orbit correction

Spin-orbit coupling is an interaction between electron-spin and the orbital angular momentum and results in the splitting of orbital energy levels. The spin-orbit correction term, ΔE_{SO}, refers to the difference between the ground state energy and the weighted average of all spin-orbit states.

In this work, in accordance with the guidelines set by the HEAT protocol, we consider only first-order spin-orbit corrections.[61] Therefore, only rad-

icals with degenerate ground states have nonzero spin-orbit corrections. For all of the species in this study, which are closed-shell hydrocarbons, the spin-orbit correction is by definition zero. Second-order spin-coupling effects would include the coupling of the ground state with the excited state of different spin through the spin-orbit operator.[1] This approximation is expected to be insignificant for the thermochemical data relevant to this paper.

1.4 Computational Details

Calculation of HF-SCF, CCSD, CCSD(T), CCSDT, DBOC, ZPE, and relativistic energy calculations come from a version of CFOUR[58] running on a local computer cluster with the LINUX operating system. For CCSDTQ calculations, the MRCC[8] chemistry program was used, coupled to CFOUR.

Chapter 2

Results

2.1 Summary of Results

The total HEAT energies for all atoms in this study, as well as their respective contributions from the above sources to the total energy, are tabulated in Table 2.1. Results are listed using the HEAT-345(Q) based method.[3] All species are closed-shell hydrocarbons, and vary in size from 5 to 17 atoms. The most important values are the HEAT total energies. These energies can be used to calculate atomization energies and standard enthalpies of formation at 0K.

The CCSDT, cc-pVQZ calculations for neopentane and isobutane were excluded from this study due to time restrictions. Each of these calculations would require several weeks or months on the local computer cluster. Because of this omission, the total HEAT energy for these two molecules must also be excluded from this work. Regardless, we choose to include neopentane and isobutane in the results in order to show the varying contributions to the total HEAT energy and atomization energy for each of these molecules.

From these total HEAT energies, we calculated the total atomization energies and tabulated the results in Table 2.2. Of the 11 species, 7 have

Table 2.1: Contributions to the HEAT total energies for the 11 species studied. All contributions are in atomic units where 2625.4976 kJ mol^{-1} = 1 E_h.

Species	E^{∞}_{HF}	$\Delta E^{\infty}_{CCSD(T)}$	ΔE_{CCSDT}	ΔE_{CCSDTQ}	ΔE_{REL}	ΔE_{ZPE}	ΔE_{DBOC}	Total
Methane	-40.217243	-0.297867	-0.000303	-0.000147	-0.016001	0.044131	0.002591	-40.484838
Ethane	-79.266953	-0.559688	-0.000328	-0.000396	-0.031977	0.073666	0.004728	-78.780950
Ethylene	-78.070966	-0.517981	-0.000167	-0.000621	-0.032069	0.050164	0.004225	-78.567415
Acetylene	-78.855702	-0.480476	0.000223	-0.001020	-0.032142	0.026213	0.003674	-77.339229
Propane	-118.318011	-0.823560	-0.000307	-0.000654	-0.047964	0.102039	0.006848	-119.081609
Propene	-117.126631	-0.781517	-0.000131	-0.000917	-0.048054	0.078514	0.006336	-117.872399
Propyne	-115.917854	-0.741684	0.000245	-0.001301	-0.048118	0.054677	0.005795	-116.648241
Allene	-115.915376	-0.741676	-0.000002	-0.001155	-0.048131	0.053867	0.005829	-116.646644
Isobutane	-157.369371	-1.089553	——[a]	-0.000940	-0.063959	0.129726	0.008955	——[a]
Isobutene	-156.181899	-1.046656	-0.000038	-0.001225	-0.064043	0.106053	0.008437	-157.179371
Neopentane	-196.420000	-1.357552	——[a]	-0.001244	-0.079960	0.157052	0.011054	——[a]

[a] denotes values that have been excluded due to computational time requirements

E_{CCSDTQ} values greater than 2 kJ mol^{-1} and 4 have values close to 3 kJ mol^{-1}. In working towards our goal of chemical accuracy (errors of <4.184 kJ mol^{-1}), these values demonstrate the importance of calculations beyond CCSDT.

For all species, the value of the ΔE_{CCSDT} correction is negative and the ΔE_{CCSDTQ} correction is positive. As noted in Ref. 1, when these two contributions are combined, the error tends to cancel almost entirely. For the species studied in this paper, all CCSDT contributions were *more negative* than the CCSDTQ results were positive, meaning that for all species the total energy was lowered from the CCSD(T) values.

Scalar relativistic effects uniformly reduce the atomization energy. The effect increases with the size of the molecule, due to the greater number of total electrons. Methane's relativistic correction is only -0.82 kJ mol^{-1}, whereas the

Table 2.2: Contributions to atomization enthalpies for the 11 species studied. All values are in kJ mol^{-1}.

Species	E^{∞}_{HF}	$\Delta E^{\infty}_{CCSD(T)}$	ΔE_{CCSDT}	ΔE_{CCSDTQ}	ΔE_{REL}	ΔE_{ZPE}	ΔE_{DBOC}	ΔE_{SO}	Total
Methane	1374.15	385.40	-0.43	0.33	-0.82	-115.87	0.41	-0.38	1642.80
Ethane	2308.57	676.16	-1.58	0.93	-1.67	-193.41	0.59	-0.76	2788.82
Ethylene	1794.12	566.66	-2.01	1.52	-1.39	-131.70	0.48	-0.76	2226.91
Acetylene	1229.05	468.19	-3.03	2.56	-1.16	-68.82	0.50	-0.76	1626.54
Propane	3246.51	972.30	-2.86	1.55	-2.49	-267.90	0.81	-1.13	3946.79
Propene	2744.16	861.92	-3.33	2.24	-2.21	-206.14	0.72	-1.13	3396.23
Propyne	2196.13	757.34	-4.31	3.25	-2.01	-143.55	0.72	-1.13	2806.43
Allene	2189.63	757.32	-3.66	2.86	-1.97	-141.43	0.63	-1.13	2802.24
Isobutane[a]	4185.26	1274.02	——[a]	2.24	-3.28	-340.60	1.06	-1.51	——[a]
Isobutene	3693.17	1161.39	-4.79	2.99	-3.03	-278.44	0.99	-1.51	4570.77
Neopentane[a]	5122.08	1581.00	——[a]	2.99	-4.07	-412.34	1.34	-1.89	——[a]

[a] denotes values that have been excluded due to computational time requirements

correction for neopentane is -4.1 kJ mol^{-1}. The zero-point energy correction is obviously negative for all species, as is the spin-orbit correction. Of possible interest is that the spin-orbit correction exists for the atomization energies but not for TAE values. Although all molecules examined in this paper are closed-shell, which as mentioned above would result in a first-order spin-orbit correction of zero, the HEAT protocol was applied to the carbon atom in order to find atomization energies, which has nonzero spin-orbit correction. Therefore, the magnitude of the spin-orbit correction to atomization energies increases in proportion to the number of carbon atoms in the molecule. The diagonal Born-Oppenheimer correction was positive for all species. This correction tends to have the greatest effect in molecules with more hydrogen atoms, with the largest correction in neopentane (1.34 kJ mol^{-1}) with 12 hydrogens.

Table 2.3: Contributions to the reaction enthalpies for the 11 species studied from elemental reactions. All contributions are in kJ mol^{-1}.

Species	E_{HF}^{∞}	$\Delta E_{CCSD(T)}^{\infty}$	ΔE_{CCSDT}	ΔE_{CCSDTQ}	ΔE_{REL}	ΔE_{ZPE}	ΔE_{DBOC}	ΔE_{SO}	Total
Methane	57.33	186.34	-1.94	1.88	0.11	50.94	0.10	-0.82	293.93
Ethane	203.59	359.90	-3.16	3.49	0.25	89.53	0.21	-1.64	652.18
Ethylene	367.23	361.99	-2.73	2.90	-0.01	53.80	0.10	-1.64	781.63
Acetylene	581.48	353.04	-1.71	1.86	-0.23	16.89	-0.14	-1.64	949.55
Propane	346.31	528.08	-4.25	5.08	0.37	125.08	0.28	-2.47	998.49
Propene	497.85	531.05	-3.78	4.39	0.11	89.29	0.15	-2.47	1116.58
Propyne	695.07	528.22	-2.80	3.38	-0.08	52.67	-0.07	2.47	1273.93
Allene	701.58	528.24	-3.45	3.77	-0.12	50.55	0.02	2.47	1278.12
Isobutane[a]	488.24	690.68	——[a]	6.60	0.47	158.82	0.32	-3.29	——[a]
Isobutene	629.52	695.90	-4.69	5.85	0.22	122.64	0.17	3.29	1446.32
Neopentane[a]	632.09	848.03	——[a]	8.06	0.55	191.61	0.34	-4.11	——[a]

[a] denotes values that have been excluded due to computational time requirements

While the atomization energies and HEAT total energies are the primary results of this paper, the elements in their standard states can also be used as elemental reference compounds. This approach is the approach favored by HEAT model chemistry. As mentioned above, the standard states for many elements are computationally impossible (specifically graphite), and as such we used carbon monoxide.

$$\mathrm{CO} + 2\,\mathrm{H_2} \longrightarrow \mathrm{CH_4} + \mathrm{O} \tag{2.1}$$

We chose CO because its experimental enthalpy of formation is not tied to that of the carbon atom[1] and is therefore a more stable reference. We used values from Ref. 1 for the HEAT values of CO and H_2 used to generate the reaction enthalpies.

In general, we can observe that similar trends are followed in both the

atomization enthalpies and the enthalpies from elemental reactions. Hartree-Fock and CCSD(T) values have contributions of similar magnitude, although the CCSD(T) correction is more likely to dominate within the elemental reaction method. Higher level corrections are also more likely to be a determining factor with this approach (for example the ΔE_{CCSDTQ} correction for neopentane is greater than 8 kJ mol^{-1}). The remaining contributions are all similar in magnitude between the two approaches, although the sign of the ZPE contribution is positive for all species in the elemental reaction approach, compared to the uniform negative contribution from the atomization approach.

2.2 Comparison with ATcT

In order to determine the accuracy of the theoretical data, we find it necessary to compare to highly accurate experimental data. Without the work of Ruscic et al. on the ATcT database[23], the theoretical data here could not be calibrated or compared against data of comparable accuracy. The ATcT, or Active Thermochemical Tables, provide a statistical analysis of various experimental and selected computational data. Unless noted otherwise, all "experimental data" used in this paper is from the ATcT network.

The goal of HEAT is to establish highly-accurate *ab initio* methods for finding relevant thermochemical data such as bond energies, heats of formation, total atomic energies, etc. In Ref. 1, the stated goal was generate data within the bounds of chemical accuracy (<1 kcal mol^{-1} = 4.184 kJ mol^{-1}). Subsequently, advances made to the HEAT protocol have allowed a new goal

of <1 kJ mol^{-1} to come within reach. As seen in Table 2.1, although the total atomic energies for all species succeed in the goal of attaining chemical accuracy, only 3 of the 11 species have errors of less than 1 kJ mol^{-1}. The largest of these errors occurs in propane and propene, both of which have errors of greater than 2 kJ mol^{-1}. A more detailed discussion of the probable sources of error may be found in the Discussion section.

The documentation of the individual contributions to the HEAT total energy are of interest to others interested in expanding the HEAT database. As pointed out in previous papers[1–3], the isodesmic reaction format is the most accurate when determining heats of formation and total atomic energies. Therefore, the inclusion of more molecular species in the literature will allow for more molecules and reactions to be compared along these lines.

Enthalpies of formation at 0 K ($\Delta_f H_0^\circ$) are calculated for all species using experimental enthalpies of formation for the atoms (carbon and hydrogen atoms) with HEAT atomization energies and the elemental reaction approach (as defined above). For the heats of formation from atomization energies (Total(I) in Table 2.4), it is of interest to note that, for all species, the calculation gives an enthalpy of formation that is *below* the ATcT estimate, or an atomization energy that is *above* the ATcT estimate. In other words, theory is overestimating the stability of the molecules relative to the constituent atoms. The same can be said of the results of the elemental reaction approach, except for isobutene, for which the theoretical heat of formation is slightly greater than the ATcT value. In addition, the errors in the reaction approach are

smaller, though the difference is rather small. For all species, the absolute error for Total(II) is smaller, except for acetylene, for which it is slightly larger.

Table 2.4: Comparison of heats of formation with ATcT values. The values for Total(I) come from atomization enthalpies and those for Total(II) come from reaction enthalpies. All values are in kJ mol^{-1}.

Species	Total(I)	Total(II)	ATcT
Methane	-67.30	-66.72	-66.59
Ethane	-69.88	-69.12	-68.16
Ethylene	59.97	60.33	61.04
Acetylene	228.28	228.25	228.66
Propane	-84.41	-83.46	-82.17
Propene	34.09	34.63	36.32
Propyne	191.83	191.98	192.84
Allene	196.02	196.17	197.60
Isobutane[a]	—	—	-106.83
Isobutene	2.99	3.72	3.39
Neopentane[a]	—	—	—
Mean absolute error	1.26	0.87	...
Mean signed error	-1.26	-0.80	...
RMS error	1.43	1.00	...
Maximum error	-2.24	-1.69	...

[a] denotes values that have been excluded due to computational time requirements

Chapter 3

Discussion

3.1 Remaining errors

It is relevant to discussion that the calculated HEAT values overestimate the stability of the molecules, resulting in enthalpies of formation that are too small and atomization energies that are too large. A full discussion of this error is beyond the scope of this paper, although we will propose some general considerations for its source.

One possible source of error comes from geometry calculations. As stated earlier, geometry optimization at the CCSD(T)/cc-pVQZ level has an intrinsic error of approximately 0.005 Å and 0.5°. As a result, the atomization energies will be smaller than expected. The molecule is not at its correct equilibrium position, and its TAE will be larger (less negative) than expected because it is not at the bottom of the potential well. Because this source of error contributes with the effect of making the atomization energy *smaller* than the true value, and all of the observed atomization energies are too *large*, we can safely discard this as a significant source of error. We wish to remind the reader at this point that this error cancellation is one of the strengths of the HEAT protocol.

Another potential source of error in this paper is the use of the HEAT-345(Q) method rather than a higher level of theory such as HEAT-456Q. We rationalize this decision to use HEAT-345(Q) for several reasons. Firstly, HEAT-345(Q) is the least costly method in the suite of possible HEAT methods. A higher level of theory than HEAT-345(Q) would have significantly increased the CPU time of many of the calculations in this paper. Furthemore, HEAT-345(Q), despite being the cheapest calculation, statistically does a better job of calculating TAEs than other methods. The interested reader is directed to Ref. 3 for an in-depth analysis of the statistical strength of the HEAT-345(Q) method.

The most probable source of the error associated with the molecules in this paper is the larger molecule size. While we can expect that the *relative error* of each species remain constant as we increase the size of the molecules, because the atomization energies themselves are growing larger, the *absolute error* will increase regardless. It is necessary due to this size extensivity of the absolute error that, even though the relative error might remain constant as molecule size is increased, the absolute error will grow proportional to molecule size. The unusually small error of 0.40 kJ mol^{-1} in the atomization energy of isobutene (and 0.33 kJ mol^{-1} from reaction enthalpies), the largest molecule for which full HEAT energy values was obtained, is likely due to fortuitous error cancellation rather than the inherent accuracy of HEAT for isobutene. A possible source of this error lies in precise determination of the heat of formation of the carbon atom which, as mentioned earlier in this work, still

carries an error bar (0.06 kJ mol^{-1}).[23] Therefore, as the number of carbon atoms increases, this error increases proportionally. A recent study examining the HEAT energy of benzene showed the same characteristic rise in absolute error, although benzene showed a rise in absolute error smaller than that of the molecules examined here.[62]

Another possible source of error comes from the lack of diffuse functions in the form of augmented basis set (i.e. aug-cc-pVXZ) on carbon for the calculation of higher level correlation effects (above the CCSD(T) level) and the calculation for the carbon atom. The inclusion of these diffuse functions are a possible venue for further research, as discussed in the next section.

Between the two methods of obtaining heats of formation, those based upon the formation of the molecule from an elemental reference compound has the higher accuracy. These results support the conclusion that the elemental reaction method provides the most accurate heats of formation. Context is required, however, for the choice of the CO reaction to form the molecule species. As mentioned above, isodesmic reactions provide the most accurate thermochemical values. Therefore, if we were to assume a reaction that better preserved the chemical environment of the molecular species, we can expect the error to decrease even further. This conclusion comes with the caveat that more and more molecular species need to have benchmark HEAT total energies calculated in order to obtain high-accuracy data. This necessity is a significant motivation in choosing the 11 species studied in this work. As stated above, the calculation of benchmark HEAT values for the homodesmotic/isodesmic

reaction hierarchy as described by Wheeler et al.[6] will allow the calculation of highly accurate thermochemical data.

3.2 Possibilities for further research

The first and most obvious avenue for further research is to finish the calculations for isobutane and neopentane at the CCSDT/cc-pVQZ level of theory. These calculations wiill require several months of wall time to complete on personal computers or a small cluster. In addition, the use of diffuse functions on the carbon atom could be explored. Although the effect, due to the highly accurate data is likely insignificant for smaller molecules (<1 kJ mol^{-1} for all molecules in Ref. 1), for larger molecules such as those studied in this paper, the effects could be more significant. Furthermore, more efficient means of calculating the anharmonic contribution to the zero point energy could be researched. The reader is reminded that the calculation of the anharmonic contribution to ZPE uses VPT2. Although this approach is based in perturbation theory, and for semi-rigid molecules provides a fairly accurate measure of the ZPE, for larger molecules it remains by far the most costly calculation in the HEAT protocol.

Chapter 4

Summary

The primary results of this work are the atomization energy calculations, which can be used to determine theoretical heats of formation. These benchmark HEAT values are useful in the homodesmotic/isodesmic reaction hierarchy as established by Wheeler et al.[6]

This work examines two main differences from the original HEAT protocol: the use of the ANO0 basis set when examining the zero point energy, and calculations involving larger molecules. The ANO0 basis set was used when calculating the anharmonic force fields in the zero-point energy calculation. In addition, this paper shows that the HEAT method can be applied to larger molecules that contain a few dozen electrons. We have shown these changes do not seriously degrade the quality of the results, because for all species the error when compared to accurate data from ATcT is smaller than what is required for chemical accuracy, <1 kcal mol^{-1}.

We can conclude that sub kJ mol^{-1} accuracy for larger molecules will require improvements in experimental heats of formation of the atoms. Furthermore, in order to continue using the HEAT protocol with progressively larger molecules, hardware will need to improve to be able to handle the

CCSDT/cc-pVQZ calculation and the calculation of anharmonic force fields for the zero-point energy. Efficient parallelization of these calculations could also help improve these calculation times.

Bibliography

1. A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez, and J. F. Stanton, J. Chem. Phys. **121**, 11599 (2004).

2. Y. J. Bomble, J. Vázquez, M. Kállay, C. Michauk, P. G. Szalay, A. G. Császár, J. Gauss, and J. F. Stanton, J. Chem. Phys. **125**, 064108 (2006).

3. M. Harding, J. Vázquez, B. Ruscic, A. Wilson, J. Gauss, and J. F. Stanton, J. Chem. Phys. **128**, 114111 (2008).

4. W. J. Hehre, L. Radom, P. V. R. Schleyer, and J. A. Pople, *Molecular Orbital Theory* (Wiley, New York, 1986), p. 298.

5. W. J. Hehre, R. Ditchfield, L. Radom, and J. A. Pople, J. Am. Chem. Soc. 1970, **92**, 4796.

6. S. E. Wheeler, K. N. Houk, P. Schleyer, and W. D. Allen, J. Am. Chem. Soc. **131**, 2547 (2009).

7. K. Bak, J. Gauss, P. Jørgensen, J. Olsen, T. Helgaker, et al., J. Chem. Phys. **114**, 6548 (2001).

8. M. Kállay and P. R. Surján, J. Chem. Phys. **115**, 2945 (2001).

9. Thom H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

10. Rick A. Kendall, Thom H. Dunning, and Robert J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

11. D. E. Woon and Thom H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

12. K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

13. J. Noga and R. J. Bartlett, J. Chem. Phys. **123**, 214105 (2005).

14. Y. J. Bomble, M. Kállay, J. Gauss, and J. F. Stanton, J. Chem. Phys. **123**, 054101 (2005).

15. M. Kállay and J. Gauss, J. Chem. Phys. **123**, 214105 (2005).

16. D. Feller, J. Chem. Phys. **96**, 6104 (1992).

17. David Feller, Kirk A. Peterson, and David A. Dixon, J. Chem. Phys. **129**, 204105 (2008).

18. T. Helgaker, W. Klopper, H. Koch, and J. Noga, J. Chem. Phys. **106**, 9639 (1997).

19. David Feller, Kirk A. Peterson, and J. Grant Hill, J. Chem. Phys. **135**, 044102 (2011).

20. J. Almlöf and P. R. Taylor, J. Chem. Phys. **86**, 4070 (1987).

21. J. Almlöf and P. R. Taylor, J. Chem. Phys. **92**, 551 (1990).

22. I. M. Mills, in *Modern Spectroscopy: Modern Research*, edited by K. N. Rao and C. W. Matthews (Academic, New York, 1972), p. 115-140.

23. B. Ruscic, R. E. Pinzon, M. L. Morton, G. von Laszevski, S. J. Bittner, S. G. Nijsure, K. A. Amin, M. Minkoff, and A. F. Wagner, *J. Phys. Chem. A*, **2004**, *108* (45), 9979-9997.

24. A. G. Császár, W. D. Allen, and H. F. Schaefer III, J. Chem. Phys. **108**, 9751 (1998).

25. J. M. L. Martin and G. de Oliveira, J. Chem. Phys. **111**, 1843 (1999).

26. A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kállay, and J. Gauss, J. Chem. Phys. **120**, 4129 (2004).

27. A. Karton, E. Rabinovich, J. M. L. Martin, and B. Ruscic, J. Chem. Phys. **125**, 144108 (2006).

28. Spectrochim. Acta 58, 599-898 (2002).

29. X. Zhang, S. P. Sander, and J. F. Stanton. *J. Phys. Chem. A*, **2012**, *116*, 10338-10343.

30. X. Zhang, S. P. Sander, A. Chaimowitz, G. B. Ellison, and J. F. Stanton. *J. Phys. Chem. A*, **2010**, *114*, 12021.

31. F. D. Rossini, J. Res. Natl. Bur. Stand. **6**, 1 (1931).

32. J. D. Cox, D. D. Wagman, and V. A. Medvedev, *CODATA Key Values for Thermochemistry* (Hemisphere, New York, 1989).

33. M. W. Chase, C. A. Davies, J. R. Downey, Jr., D. J. Frurip, R. A. McDonald, and A. N. Syverud, J. Phys. Chem. Ref. Data **14**, 1 (1985).

34. L. V. Gurvich, I. V. Veyts, and C. B. Alcock, *Thermodynamic Properties of Individual Substances* (Hemisphere, New York, 1989), Vol. 1, Pts. 1 and 2; *ibid.*1991), Vol. , Pts. 1 and 2; *ibid.*1996), Vol., Pts. 1 and 2.

35. B. A. Flowers, P. G. Szalay, J. F. Stanton, M. Kállay, J. Gauss, and A. G. Császár. *J. Phys. Chem. A*, **2004**, *108* (15) 3195.

36. A. D. Boese, M. Oren, O. Atasolyu, J. M. L. Martin, M. Kállay, and J. Gauss, J. Chem. Phys. **120**, 4129 (2004).

37. D. Feller, J. Chem. Phys. **111**, 4373 (1999).

38. D. Feller and J. A. Sordo, J. Chem. Phys. **113**, 485 (2000).

39. D. Feller, J. Chem. Phys. **96**, 6104 (1992).

40. M. Kállay and J. Gauss, J. Chem. Phys. **120**, 6841 (2004).

41. S. Hirata and R. J. Bartlett, Chem. Phys. Lett. **321**, 216 (2000).

42. J. Olsen, J. Chem. Phys. **113**, 7140 (2000).

43. M. Kállay and P. R. Surján, J. Chem. Phys. **115**, 2945 (2001).

44. M. Kállay and P. R. Surján, J. Chem. Phys. **113**, 1359 (2000).

45. S. Hirata. *J. Phys. Chem. A*, **2003**, *107*, 9887.

46. T. A. Ruden, T. Helgaker, P. Jørgensen, and J. Olsen, Chem. Phys. Lett. **371**, 62 (2003).

47. V. Barone, J. Chem. Phys. **120**, 3059 (2004).

48. V. Barone and C. Minishino, J. Mol. Struct.: THEOCHEM **330**, 365 (1995).

49. M. S. Schuurman, S. R. Muir, W. D. Allen, and H. F. Schaefer, J. Chem. Phys. **120**, 11586 (2004).

50. Edward F. Valeev and C. David Sherrill, J. Chem. Phys. **118**, 9 3921-3927 (2003).

51. H. Sellers and P. Pulay, Chem. Phys. Lett. **103**, 463 (1984).

52. N. C. Handy, Y. Yamaguchi, and H. F. Schaefer, J. Chem. Phys. **84**, 4481 (1986).

53. N. C. Handy and A. M. Lee, Chem. Phys. Lett. **252**, 425 (1996).

54. W. Kutzelnigg, Mol. Phys. **90**, 909 (1997).

55. D. W. Schwenke, J. Phys. Chem. A **105**, 2352 (2001).

56. O. L. Polyansky, A. G. Császár, S. V. Shirin, N. F. Zobov, P. Barletta, J. Tennyson, D. W. Schwenke, and P. J. Knowles, Science **299**, 539 (2003).

57. J. Gauss, A. Tajti, M. Kállay, J. F. Stanton, and P. G. Szalay, J. Chem. Phys. **125**, 144111 (2006).

58. CFOUR, coupled-cluster techniques for computational chemistry, a quantum-chemical program package by J. F. Stanton, J. Gauss, M. E. Harding, P. G. Szalay with contributions from A. A. Auer, R. J. Bartlett, U. Benedikt, C. Berger, D. E. Bernholdt, Y. J. Bomble, L. Cheng, O. Christiansen, M. Heckert, O. Heun, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, K. Klein, W. J. Lauderdale, D. A. Matthews, T. Metzroth, D. P. ONeill, D. R. Price, E. Prochnow, K. Ruud, F. Schiffmann, W. Schwalbach, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, J. D. Watts and the integral packages MOLECULE (J. Almlöf and P. R. Taylor), PROPS (P. R. Taylor), ABACUS (T. Helgaker, H. J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version; see http://www.cfour.de.

59. W. Klopper, J. Comput. Chem. **18**, 20 (1997).

60. C. Michauk and J. Gauss, J. Chem. Phys. **127**, 044106 (2007).

61. J. L. Tilson, W. C. Ermler, and R. M. Pitzer, Comput. Phys. Commun. **128**, 128 (2000).

62. M. E. Harding, J. Vázquez, J. Gauss, J. F. Stanton and M. Kállay. J. Chem. Phys. **135**, 044513 (2011).

63. J. M. L. Martin, P. R. Taylor, and T. J. Lee, Chem. Phys. Lett. **275**, 414 (1997).

64. L. McCaslin and John F. Stanton. Molecular Physics (2013): Calculation of fundamental frequencies for small polyatomic molecules: a comparison between correlation consistent and atomic natural orbital basis sets, Molecular Physics: An International Journal at the Interface Between Chemistry and Physics, DOI: 10.1080/00268976.2013.811303

65. A. Karton, P. R. Taylor, and J. M. L. Martin, J. Chem. Phys. **127**, 064104 (2007).